"GAUSSIANIZATION" METHOD FOR IDENTIFICATION OF MEMORYLESS NONLINEAR AUDIO SYSTEMS

I. Marrakchi-Mezghani[(1)]*,G. Mahé*[(2)]*, M. Jaïdane-Saïdane*[(1)]*, S. Djaziri-Larbi*[(1)]*, M. Turki-Hadj Alouane*[(1)]

[(1)] Unité Signaux et Systèmes, Ecole Nationale d'Ingénieurs de Tunis, Tunisia
email: U2S@enit.rnu.tn, {m.turki, sonia.larbi, meriem.jaidane}@enit.rnu.tn
[(2)] Centre de Recherche en Informatique de Paris 5 (CRIP5), Université Paris Descartes, France
email: gael.mahe@math-info.univ-paris5.fr

ABSTRACT

Identification and compensation purposes of nonlinear systems are of interest for many audio processing applications. The analysis of systems under test must be done through realistic audio inputs in order to capture different aspects of the nonlinearity. However, the Gaussianity of the tested signal, is a desirable factor because it guarantees easy implementation and good performances for the nonlinearity identification process. In this paper, we show at a first stage, the importance of input Gaussianity for the identification of memoryless nonlinear systems. At a second stage, we propose an algorithm that makes the speech signals Gaussian. The proposed "Gaussianization" algorithm is based on the embedding of an imperceptible signal in the speech signal, to force it to be Gaussian. As expected, the performances of the optimal identification of a polynomial nonlinearity are much better with the Gaussianized input than with the original one. Moreover, these performances exhibit a robustness similar to the Gaussian input case.

1. INTRODUCTION

Identification or compensation (predistorsion/equalization) purposes of acoustic nonlinear systems (transducers systems within loudspeakers and microphones) are of interest for many audio processing applications (hands-free telephone systems, visioconference systems, high quality loudspeaker systems,...).

Usually nonlinear distortions are modelled either by memoryless nonlinearities or by nonlinear systems with memory. Various structures such as neural networks [1], Volterra filters [2], Wiener/Hammerstein systems, polynomial structures, etc. were tested.

Optimal approaches for the identification or the characterization of these nonlinear systems are generally made with "academic" inputs such as one or two tones, or with white Gaussian inputs. Whereas in practice, these audio systems are driven by audio signals such as speech that are more complex (non stationary, correlated, with Laplacian or Generalized Gaussian distribution function). To capture different aspects of the nonlinear distortions as with realistic signals, synthetic signals with required properties are designed for this use. For example, USASI data that are stationary correlated noise with the same spectrum as speech, was used to test acoustic echo canceller.

In this paper, we show the importance of input Gaussianity for the identification of memoryless nonlinear systems. To exploit the features of a realistic speech signal to identify the nonlinearity, we propose an algorithm that makes the speech signals "Gaussian".

The proposed "Gaussianization" is based on the embedding of an imperceptible information in the audio signal, which forces the audio signal to be Gaussian. The proposed approach has the same viewpoint as the audio stationnarization approach used to enhance the performances of a real-time adaptive echo canceller system [3, 4].

This paper is organized as follows : we show in section 2 the importance of the Gaussian property for nonlinear system identification. A "Gaussianization" method for audio signals based on the embedding of an imperceptible signal to force the speech input to be Gaussian is proposed in section 3. Simulation results with original and "gaussianized" inputs are presented and discussed in section 4.

2. IMPORTANCE OF INPUT GAUSSIANITY FOR IDENTIFICATION OF MEMORYLESS NONLINEAR AUDIO SYSTEMS

In this section, we analyze the influence of the input Probability Density Function (PDF) on the optimal identification -in the mean square sense- of a nonlinear audio system using a memoryless polynomial model. We prove that the identification performances are all the more sensitive as the nonlinearity order increases. We can quantify then the expected identification performances when gaussianized speech inputs are used instead of speech signals (assumed to be Laplacian).

2.1 Optimal identification and input PDF influence

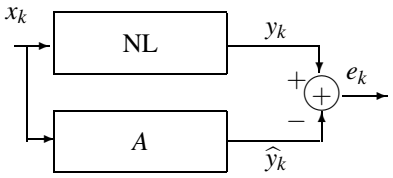

FIG. 1 – Nonlinear system identification.

As shown in Fig. 1, x_k denotes the input of the nonlinear (NL) audio system, y_k is its output and $\widehat{y}_k = A^T X_k$ corresponds to the output estimate. $A = (a_1, a_2, ..., a_p)^T$ is the parameter vector to be optimized and $X_k = (x_k, x_k^2, ..., x_k^p)^T$ is the observation vector of the p-order polynomial model.

The optimal vector A^{opt} that minimizes the Mean Squared Error (MSE) $E[(y_k - \widehat{y}_k)^2]$ is given by :

$$R_x A^{opt} = r_{Xy},$$

where $R_x = E[X_k X_k^T]$ and $r_{Xy} = E[X_k y_k]$.

The symmetric matrix R_x is defined by :

$$\begin{pmatrix} m_2 & m_3 & . & . & m_{p+1} \\ m_3 & . & . & m_{p+1} & m_{p+2} \\ . & . & m_{p+1} & m_{p+2} & . \\ . & m_{p+1} & m_{p+2} & . & . \\ m_{p+1} & m_{p+2} & . & . & m_{2p} \end{pmatrix}, \quad (1)$$

where $m_i = E[x_k^i]$ is the i^{th} order moment.

The robustness of the identification performances are closely related to the conditioning of the matrix R_x which depends on the PDF of the input signal x_k. Typically the conditioning of a symmetric matrix R_x is evaluated through its logarithmic condition number [6] :

$$K(R) = Log_{10}\left(\frac{|\lambda_{max}|}{|\lambda_{min}|}\right), \quad (2)$$

where λ_{max} and λ_{min} are respectively the largest and the smallest eigenvalues of the matrix R_x.

Audio signals such as speech signals are assumed to be Laplacian whereas music signals are rather Gaussian [5]. Therefore, we propose in the following, a comparative study of $K(R)$ corresponding to a p-order polynomial nonlinearity model and an input with Gaussian or Laplacian distribution.

Applying the Price theorem [7] on a zero mean Gaussian process x^g, we may deduce all the higher order moments through $\sigma_{x^g}^2 = E[(x^g)^2]$:

$$\begin{aligned} m_{2n+1}^g &= E[(x^g)^{2n+1}] = 0, \\ m_{2n}^g &= \frac{(2n-1)!}{2^{n-1}(n-1)!}\sigma_{x^g}^{2n}, \end{aligned} \quad (3)$$

where $n > 0$.

Similarly, for a zero mean Laplacian process x^l :

$$\begin{aligned} m_{2n+1}^l &= E[(x^l)^{2n+1}] = 0, \\ m_{2n}^l &= \frac{(2n)!}{2^n}\sigma_{x^l}^{2n}, \end{aligned} \quad (4)$$

where $\sigma_{x^l}^2 = E[(x^l)^2]$. For a nonlinear model of order $p = 3$, for example :

$$R_{x^g} = \begin{pmatrix} \sigma_{x^g}^2 & 0 & 3\sigma_{x^g}^4 \\ 0 & 3\sigma_{x^g}^4 & 0 \\ 3\sigma_{x^g}^4 & 0 & 15\sigma_{x^g}^6 \end{pmatrix} \quad (5)$$

and

$$R_{x^l} = \begin{pmatrix} \sigma_{x^l}^2 & 0 & 6\sigma_{x^l}^4 \\ 0 & 6\sigma_{x^l}^4 & 0 \\ 6\sigma_{x^l}^4 & 0 & 90\sigma_{x^l}^6 \end{pmatrix}. \quad (6)$$

Assuming that x^g and x^l are unit variance processes, for $p = 3$, R_{x^g} is better conditioned than R_{x^l} since $K(R_{x^g}) = 1.6$ and $K(R_{x^l}) = 2.18$.

This important result can be generalized for any nonlinearity order p. Indeed, as depicted on Fig. 2, the matrix R_{x^g} is better conditioned than R_{x^l} for all considered values of p. However, as expected, this shows that in the case of speech signal the matrix is ill-conditioned. The condition numbers of the estimated matrices $\widehat{R}_{x^s}$ for speech were computed for $N = 10000$ samples.

Thus, we expect that the optimal identification of a nonlinear system by a memoryless polynomial model will have better performances for Gaussian input than for Laplacian input.

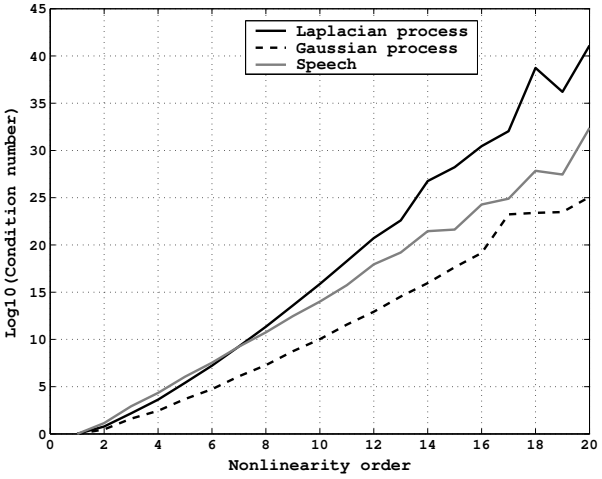

FIG. 2 – Condition number evolution of R_{x^g} (Gaussian process), R_{x^l} (Laplacian process) and $\widehat{R}_{x^s}$ (speech signals).

2.2 Variability of identification performances

The nonlinear system NL is supposed to be a memoryless polynomial system of order Q (where $Q \geq p$). In the following, we evaluate the identification performance with respect to the nonlinearity order, for white Gaussian, Laplacian and speech inputs.

2.2.1 Exact modelling

When $Q = p$, we evaluate (Fig. 3) the mean relative estimation error :

$$e_A = \frac{\sqrt{\frac{1}{M}\sum_{i=1}^{M} e_i^2}}{\|F\|}, \quad (7)$$

where F is the parameter vector of the nonlinear system (NL), $e_i = \|F - A_i^{opt}\|$ denotes the estimation error for the i^{th} frame of length $L = 256$ samples, A_i^{opt} is the parameter vector obtained through the optimal identification for the i^{th} frame and M is the number of frames.

According to the previous results on matrix conditioning (Fig. 2), the mean estimation error :

- increases with the nonlinearity order
- is smaller for a white Gaussian signal than for white Laplacian or speech signals.

In the exact modelling case, the expected contribution of "Gaussianization" is limited to a small range of nonlinearity orders between $12 \leq p \leq 17$ (the error is less than 1% until $p = 12$ for Laplacian or speech signals and near to 100% from $p = 17$ for white Gaussian input).

2.2.2 Undermodelling

In the real undermodelling context, the identification performance is measured by the Signal to Error Ratio (SER) defined as :

$$SER(k) = 10\log_{10}\left(\frac{E[y_k^2]}{E[e_k^2]}\right), \quad (8)$$

where y_k is the system output and $e_k = y_k - (A^{opt})^T X_k$ is the identification error.

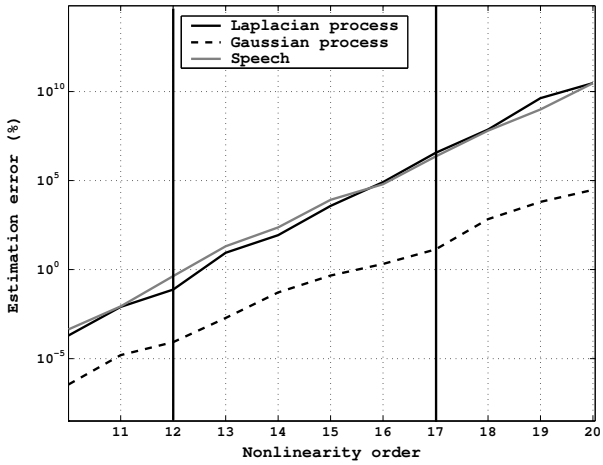

FIG. 3 – The mean relative estimation error for Gaussian, Laplacian and speech signals (exact modelling case $Q = p$).

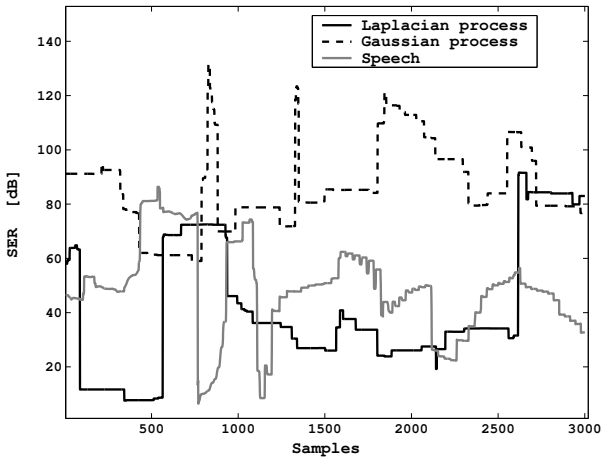

FIG. 4 – *SER* time evolution for white Gaussian, Laplacian and speech inputs (undermodelling case $Q = 17$, $p = 15$).

We depict on Fig. 4 the *SER* computed over sliding frames of length $L = 256$ samples. For these reported results, the optimal nonlinearity parameter vector is computed over disjoint frames of length L. The considered undermodelling case is for $Q = 17$ and $p = 15$.

Some conclusions about the expected performances with "gaussianized" inputs can be deduced :

- performance enhancement : the *SER* with white Gaussian input is always higher than for white Laplacian input or speech.
- robustness : even for stationary inputs (Gaussian or Laplacian), the *SER* is variable due to the variability of the conditioning number through its realizations. However, the most robust *SER* performances are obtained for Gaussian input.

3. SPEECH "GAUSSIANIZATION" METHOD

The proposed "Gaussianization" method is based on slight changes of speech sample values (that preserve the inaudibility) so that the obtained PDF matches a normal distribution.

3.1 Gaussianization principles and algorithm

A "Gaussianization" method was proposed in [8], in a context of adaptive filtering. The principle of this method is to (i) use the empirical density transformation, to obtain an uniform density ; (ii) transform to a gaussian distribution, using the inverse of the gaussian cumulative distribution function. We propose a direct and slight transformation from the empirical distribution, which is laplacian for speech signals, to a gaussian distribution. Denoting by x_k the speech signal, the different steps for its Gaussianization, which is done for disjoint frames of length $N = 10000$ samples, are :

1. normalizing each frame of the signal so that we obtain a unit variance and zero mean sequence.
2. arranging in an ascending order the samples of the normalized sequence $X = \{x_1, x_2, ..., x_N\}$. We get the corresponding ordered sequence $Z = \{z_1, z_2, ..., z_N\}$, which has the empirical cumulative distribution function :

$$F_Z^{emp}(z_k) = P[Z \leq z_k] = \frac{k}{N}\ ,\ k = 1 : N. \qquad (9)$$

The values of the target normal cumulative distribution function for the sequence Z are :

$$F_Z^{th}(z_k) = 1 - \frac{1}{2} erfc\left(\frac{z_k}{\sqrt{2}}\right), \qquad (10)$$

where $erfc$ is the complementary error function.

3. for $k = 1 : N$, adding a small value g_k to z_k in order to get the value z_k^g so that :

$$F_Z^{emp}(z_k^g) = F_Z^{th}(z_k^g),$$

as illustrated in Fig. 5.

4. rearranging the samples of the sequence $Z^g = \{z_1^g, z_2^g, ..., z_N^g\}$ in the initial order (the order of X) to get the "Gaussianized" sequence $X^g = \{x_1^g, x_2^g, ..., x_N^g\}$,
5. denormalizing the obtained sequence X^g.

Hence, the "Gaussianized" signal can be written as :

$$x_k^g = x_k + g_k,$$

where g_k represents the additive "Gaussianization" noise. The proposed procedure can be compared to a watermarking process where the additive watermark signal is g_k and the watermarked one is x_k^g.

3.2 Inaudibility and enhanced "Gaussianization"

The noise introduced by the Gaussianization process is clearly audible. This can be explained by the fact that the PDF of speech is much higher than the Gaussian one around zero. Consequently, the Gaussianization noise added to low level segments of speech induces a low signal to noise ratio in these areas. Such phenomenon is particularly noticeable for :

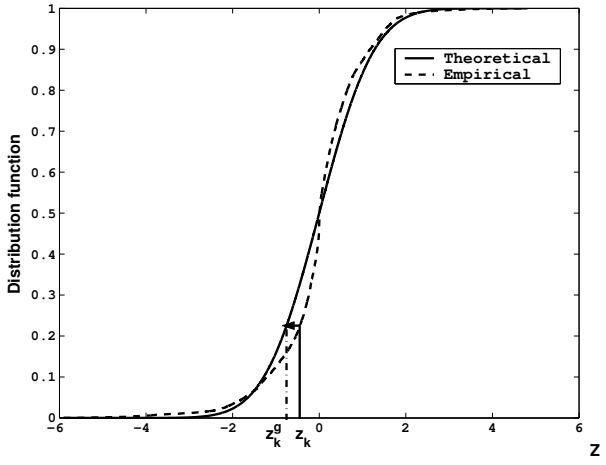

FIG. 5 – Empirical and target distribution functions of a speech signal of length $N = 10000$ samples.

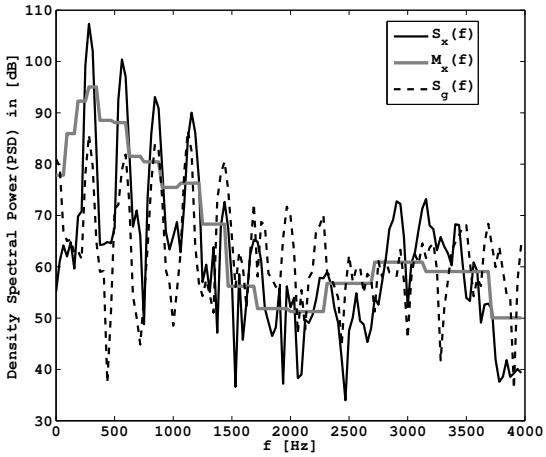

FIG. 6 – PSDs of speech signal $S_x(f)$, Gaussianization noise $S_g(f)$ and masking threshold $M_x(f)$.

– silence areas where the signal values are very small or null (values in $\{-q,0,q\}$ or in $\{-q/2,q/2\}$ where q denotes the quantization step size generally used by speech codecs)
– unvoiced areas where the zero crossing rate is high and consequently the PDF is to high near zero.

In audio applications, the transmitted sound must be perceptually unchanged compared to the original one. Hence, in order to reduce the Gaussianization noise, we propose to exclude the silence and the unvoiced segments from the sets of samples to be Gaussianized. However, the Gaussianization noise is still audible.

Fig. 6 shows the Power Spectral Density (PSD) of the Gaussianisation noise for a speech signal frame of duration 32 ms, compared to the frequency masking threshold of the signal (computed with the MPEG1 auditory model). As expected, the PSD of the Gaussianization noise is above the masking curve. Thus, we propose in the following an improved version of Gaussianization algorithm.

3.3 The proposed Gaussianization method

It can be noticed, in Fig. 6, that the PSD of the Gaussianization noise is rather parallel to the masking threshold of the audio signal. Thus, we propose to achieve the frequency masking through an iterative limitation of the maximum amplitude of g_k, according to the following steps :

– getting the ordered sequence Z with mean value 0 and variance 1, as described in steps 1 and 2 of the basic algorithm presented in subsection 3.1
– fixing the target variance of the Gaussianization noise

$$\sigma_g^{target} = \lambda,$$

where λ is the attenuation factor ($0 \leq \lambda \leq 1$).
– initializing the maximum allowed amplitude of g_k, $|g_k|_{max} = \sqrt{3}\sigma_g^{target}$. This initial value is based on the hypothesis that g will have a uniform distribution
– repeating step 3 of the basic algorithm under the constraint $|g_k| \leq |g_k|_{max}$, until $\sigma_g = \sigma_g^{target}$. The value of $|g_k|_{max}$ is adjusted at each iteration according to a dichotomic process based on the value of σ_g.

Informal tests indicate that for $\lambda_{dB} = 20Log_{10}(\lambda) = -20$ dB, the perceptual quality of the original signal and its Gaussianized version is piecewise the same. This results is illustrated in Fig. 7 where we depict the PSD of a speech signal, of the Gaussianization noise under inaudibility constraint and the frequency masking threshold $M_x(f)$. One can notice that, under constraint, the PSD of the Gaussianization noise is quietly under the masking threshold.

In the following section, we will show the enhancement of "Gaussianization" under inaudibility constraint in the identification performances for nonlinear systems.

4. IDENTIFICATION PERFORMANCES AND ROBUSTNESS FOR "GAUSSIANIZED" SPEECH

The parameter settings chosen for all reported simulations are :

– a speech signal sampled at 8 kHz,
– "Gaussianization" of disjoint frames of length $N = 10000$ samples,
– an attenuation factor $\lambda_{dB} = -20$ dB.

4.1 Improvement of the conditioning

In Fig. 8, we plotted the condition numbers for a white Gaussian noise, a speech signal and its "Gaussianized" versions (with and without inaudibility constraint) for different nonlinearity orders.

Then, we can conclude that the proposed "Gaussianization" method improves the conditioning of the input matrix R_x.

4.2 Performance evaluation

Since the matrix R_x is better conditioned for "Gaussianized" signal than for its original version, the identification performances should be better using the proposed "Gaussianization" method.

A polynomial memoryless system of order $Q = 17$ was simulated and identified through a model of order $p = 15$, over speech frames of length $L = 256$ (32 ms). The *SER*,

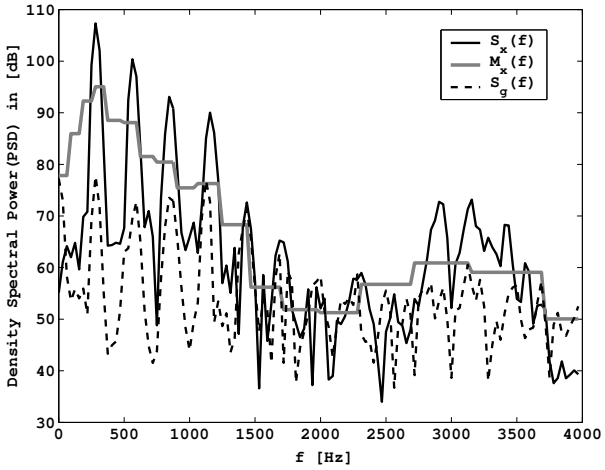

FIG. 7 – PSDs of a speech signal $S_x(f)$, of the Gaussianization noise $S_g(f)$ (under constraint $\lambda_{dB} = -20$ dB) and the masking threshold $M_x(f)$.

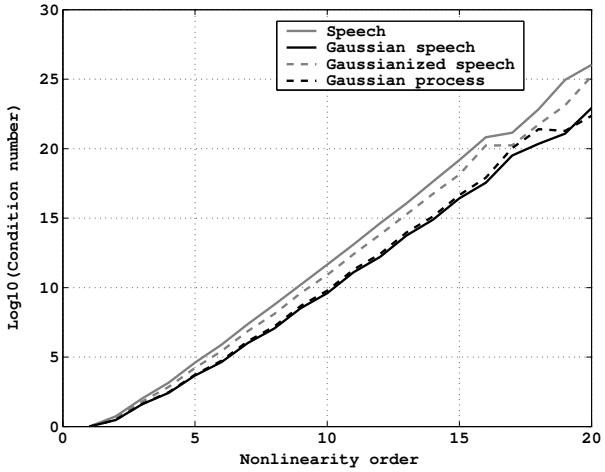

FIG. 8 – Condition number of the matrix R_x for original speech signal, Gaussian version without constraint, Gaussianized version under constraint and for a Gaussian process.

displayed in Fig. 9, confirms the good identification performance achieved through the "Gaussianization", even with the inaudibility constraint. With this constraint, the proposed method performs an average gain of 25 dB.

5. CONCLUSION

In this paper we have investigated the impact of signal gaussianity and speech "Gaussianization" for memoryless nonlinear audio systems identification/characterization. Since speech signals are rather Laplacian, we have proposed a specific method of "Gaussianization" by embedding an imperceptible information to force the speech input to be Gaussian. Polynomial systems are tested and very promising results are obtained as the nonlinearity order increases. In particular, the identification is more robust with "Gaussianized" speech than with original speech.

Now, an identification and compensation study shall be carried out for nonlinear systems with memory and for audio signals inputs (wide-band speech and music) to further emphasize the interest of the proposed procedure in a most general case.

** This work is related to the project WaRRIS (Watermarking Réflexif pour le Renforcement des Images et des Sons) supported by the French National Research Agency (ANR 2006-2009).*

REFERENCES

[1] C. Botoca and G. Buruda, *"Neural Networks intelligent tools for telecommunications problems"*, *IEEE Trans. on Electronics and Communications*, vol. 48(62), 2003.

[2] S. Stenger and R. Rabenstein, *"Adaptive Volterra filters for nonlinear acoustic echo cancellation"*, Proc. NSIP'99, Turkey, 1999.

[3] S. Larbi and M. Jaïdane, "Audio watermarking : a way to stationnarize audio signals", *IEEE Trans. on Signal Processing*, vol. 53(2) :816-823, 2005.

[4] I. Marrakchi, M. Turki, S. Larbi, M. Jaïdane and G. Mahé, "Speech processing in the watermarked domain : application in adaptive acoustic echo cancellation ", *EUSIPCO*, Italy, 2006.

[5] S. Gazor and W. Zhang, *"Speech probability distribution"*, IEEE Signal Processing Letters, vol. 10(7), July 2003.

[6] G. H. Golub, F. Van Loan. *Matrix Computations*, 3^{rd} Edition, Johns Hopkins University Press, 1996.

[7] M. Hill, *"Probability, random variables and stochastic processes : Price's theorem and joint moments"*, 2^{nd} Edition, pp 226-228, NewYork, 1984.

[8] P. Kidmose, *"Adaptive filtering for non-gaussian processes"*, *Proc. ICASSP*, vol. 1 : 424-427, Turkey 2000.

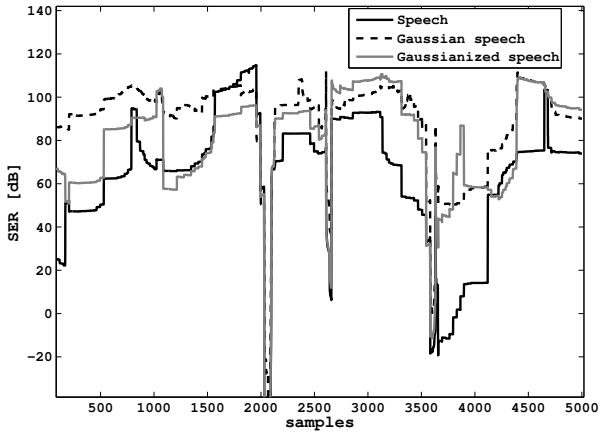

FIG. 9 – *SER* for speech signal, Gaussian version (without constraint) and "Gaussianized" version (under constraint).